

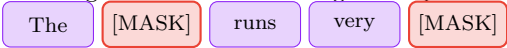
Function: Initialization

Conditions first token generation

Function: Termination

Signals sequence completion

MASK Token - Bidirectional Learning



Comparative Analysis

Function: Prediction

Aspect	SOS Token	EOS Token	MASK Token
Position	Start only	End only	Any position
Context	No context	Left context	Bidirectional
Training	Autoregressive	Autoregressive	MLM
Purpose	Initialization	Termination	Prediction
In Output	Removed	Optional	Replaced
Models	GPT, T5	GPT, T5	BERT, RoBERTa

Usage Patterns in Different Architectures

GPT-style Models

Autoregressive Generation



Applications:

Text generation, Translation,  
Dialogue systems

BERT-style Models

Bidirectional Understanding



Applications:

Language understanding,  
Classification, QA